



State of MediaWiki

Brion Vibber
Wikimedia Foundation, Inc.

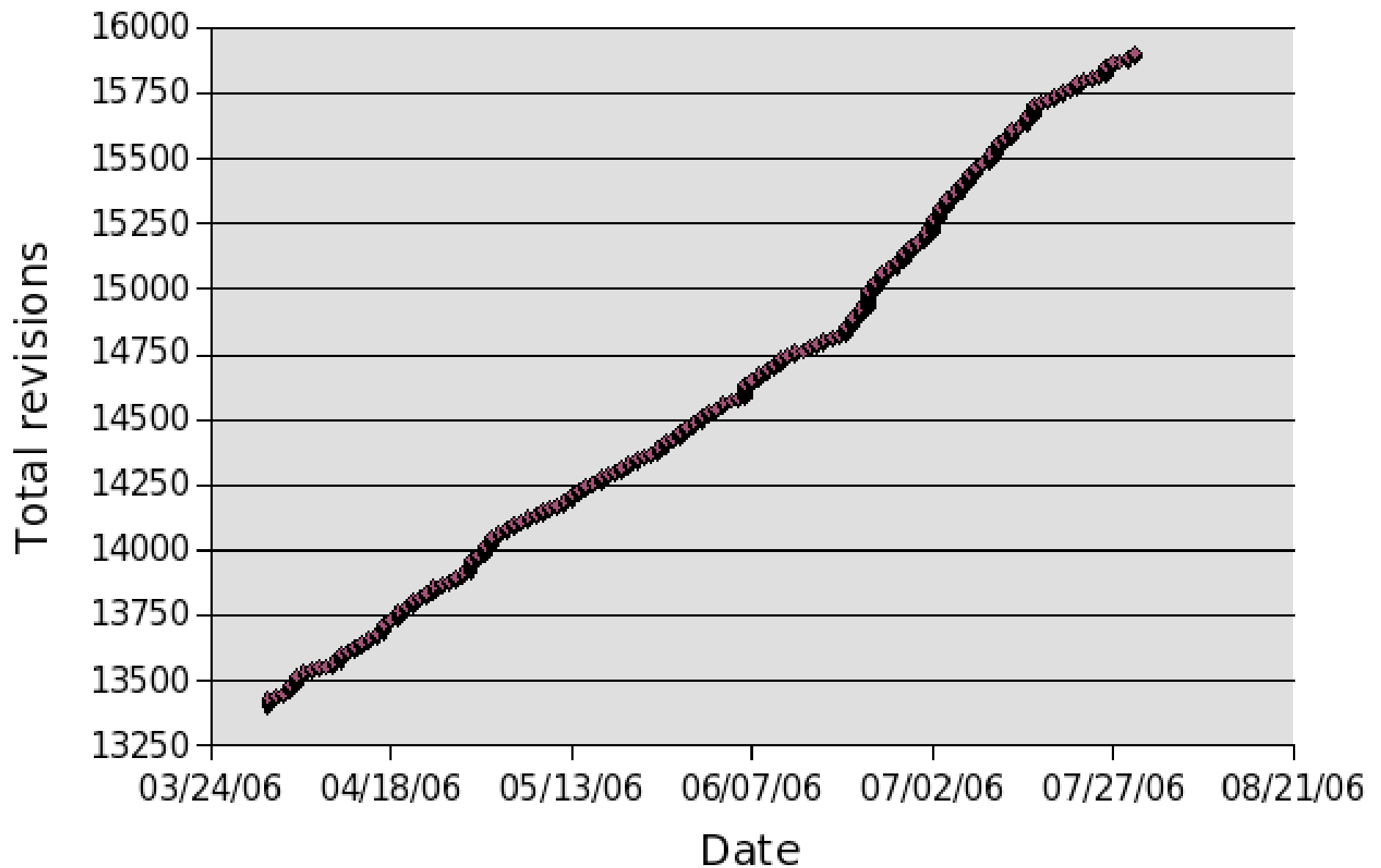
August 2, 2006

Wikimania Hacking Days
Cambridge, MA

MediaWiki

- Good
 - does most of what we need
 - scales OK
- Bad
 - Still very bad performance in some places
 - Parser too irregular

Subversion commits



Performance

- Overall performance has been good this year
 - Splitting English database
 - Faster image upload server
 - Better load balancing, caching
 - Keeping number of machines at or ahead of the growth curve
 - Lots of profiling and targeted software work by Tim and Domas

Reliability

- Good enough for free, but not great:
 - Two cluster restarts from power failures early this year
 - Occasional database explosions
 - Occasional panics from death of critical machines

Reliability possibilities

- Move from installation scripts + ad hoc updates to ready-to-run system images
- Consider running critical systems in a virtual environment for easy transplantation to a new physical box (Xen, VMWare)
- Ensure a standby for large file servers
- Clarify and improve data backup and recovery procedures

Full-text search

- Search performance problems largely eliminated by improving load balancing and adding more machines.
- Current Lucene search server searches OK, but doesn't update quickly.
- Could work to improve it, or replace it...
- Alternative engines? (sphinx, etc)
- Better implementations? (open source)

Cool things to come?

- Semantic MediaWiki!!!! I I I one
- Xapian search? (xapian.org)
- Wikidata?
- LiquidThreads?
- Better video, PDF, DjVu integration?
- External auth: OpenID / YADIS / A-Select?
- WYSIWYG?